On Hypothesis Testing in RAIM Algorithms: Generalized Likelihood Ratio Test, Solution Separation Test and a Possible Alternative

El-Mowafy A.[1], Imparato D.[1], Rizos C.[2], Wang J.[2], and Wang K.[1]

[1]School of Earth and Planetary Sciences, Curtin University, Bentley, WA 6845, Australia
[2]School of Civil and Environmental Engineering, UNSW, Kensington, NSW 2052, Australia

March 23, 2019

Abstract

Integrity for GNSS-based navigation can be monitored at user level by means of RAIM (Receiver Autonomous Integrity Monitoring) algorithms. Most of these algorithms are based on statistical tests that are able to detect and identify outliers or other anomalies in the measurements, and then either exclude suspected measurements from the position solution or forward a warning to the user. In this paper the two statistical tests most commonly used in RAIM algorithms, the Generalized Likelihood Ratio (GLR) test and the Solution Separation (SS) test, are compared. The main differences between the two tests are pointed out, in general statistical terms and in view of their use in integrity monitoring. As both tests are found not optimal for integrity monitoring, a new test is proposed that targets only the faults that represent a threat to the integrity. Simulation results are shown to substantiate the theoretical findings, and confirm the effectiveness of the new testing procedure.

1 Introduction

In the context of increasing employment of GNSS in safety applications of Intelligent Transportation Systems (ITS), one of the most stringent requirements is the capability of providing error bounds to the position solution with an extremely high confidence level, thereby guaranteeing integrity. RAIM (Receiver Autonomous Integrity Monitoring) algorithms exploit the redundancy of satellite measurements to guarantee integrity, in principle without any aid from external augmentation systems (e.g. SBAS, GBAS). Most RAIM algorithms are based on tests that are able to detect and identify possible anomalies in the observations, and eventually exclude suspected measurements from the position solution. One could also forward a warning (alert) in order to maintain the required integrity level.

Different detection/identification tests have been proposed over the years. This paper focuses on the two most commonly used tests: the standard Generalized Likelihood Ratio (GLR) test and the Solution Separation (SS) test. The two tests are compared and differences are pointed out, first from a general statistical point of view and then from an integrity focused point of view.

The GLR test is derived from the Neyman-Pearson lemma [15] and was developed for geodetic applications by the Delft School in [2, 3, 18, 19] where it was employed as a GLR and Detection, Identification and Adaptation (DIA) algorithm. The SS statistic was first presented in [16] and in [7], and is now being adopted in ARAIM algorithms (see for instance [6, 10]). An attempt to point out their differences and reconcile the two theories has been made in [4, 5].

The SS method uses as a test statistic the estimator of the bias that would be introduced in the position solution by a specific anomaly in the observations. In this way the test focuses only on the detection of faults that have sensible impact on the position solution. The GLR approach adopts as a test statistic the estimator of the bias which may possibly be present in the range measurements, therefore it applies to any general type of anomaly.

As an alternative method to the GLR and SS approach, in this study, a new test is proposed in this paper that targets only the faults that represent a threat to the integrity. This new test is still based on the

application of the Neumann-Pearson theorem, but to differently defined null and alternative hypotheses. In the proposed approach, the alternative hypothesis is defined by the presence of biases in the observations that constitute a hazard for the navigation, i.e. they exceed a certain allowable size.

Simulations of a multi-constellation Single Point Positioning (SPP) scenario with artificially injected biases in pairs of observations were run to compare the different tests. Note that the tested scenarios in the simulations only aim to show examples of the results produced by all the three methods for demonstration purpose. More complicated situations could appear in real applications, which could be interesting for further investigations, but is not attemped to be discussed in the framework of this study.

The paper is organized as follows: in Sections 2 the underlying GNSS positioning model and the statistical hypotheses to be tested are introduced. Section 3 and 4 present the GLR and the SS tests respectively. In Section 5 the two tests are compared. The most important mathematical relationships between the test statistics employed in the two procedures and their practical differences are pointed out. In Section 6 the proposed test addressing only hazardous biases is introduced. Section 7 describes the simulations set-up and results, and a discussion of the simulations is given in Section 8. Conclusions and recommendations are presented in Section 9.

2 Mathematical model

GNSS observations model can be approximated by a linear model expressed as [14]:

$$\underline{y} = Ax + \underline{e} \tag{1}$$

with

$$\underline{e} \sim N(0, Q_y)$$

where $\underline{y}$ is the vector of measurements of size m, x is the unknown position vector of size $n \leq m$, A is the full rank $m \times n$ geometry matrix and $\underline{e}$ is the measurement noise vector (of size m). The Gaussian model is most commonly adopted to describe the error distribution. Underlines are used in this paper to indicate random variables.

The above system represents the state of standard operations, i. e. the case in which the system is working properly without any fault. This state is considered as the null hypothesis $\mathcal{H}_0$. In case of a fault affecting the system, the alternative hypothesis $\mathcal{H}_a$, the linear model can be written in the following way [9]:

$$\begin{array}{ll} \mathcal{H}_0: & \underline{y} = Ax + \underline{e} \\ \mathcal{H}_a: & \underline{y} = Ax + C\nabla + \underline{e} \end{array} \tag{2}$$

where C is a $m \times q$ matrix that specifies the location of the bias in the measurements. ∇ is a vector of size q that contains the magnitudes of the bias in each of the q dimensions of the anomaly. The product $C\nabla$ is normally referred to as ∇y since it represents the full measurements bias. This model can handle multi-dimensional faults, since q can assume any value from 1 to $m-n$. The alternative hypothesis increases and the model becomes weaker as the number of the anomalies (q) increases. When we are interested in detecting a fault of a particular signature, it makes sense to design an alternative hypothesis (with the corresponding C) that addresses that particular signature.

3 GLR test statistic

The GLR test statistic for the model described in the previous section for testing the hypotheses in (2) reads [14]:

$$\underline{T}_q = \underline{\hat{e}}_0^T Q_y^{-1} C (C^T Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C)^{-1} C^T Q_y^{-1} \underline{\hat{e}}_0 \tag{3}$$

where $\underline{\hat{e}}_0 = \underline{y} - A\underline{\hat{x}}_0$ is the vector of observation residuals computed considering the null hypothesis to be true. $\underline{\hat{x}}_0$ is the position estimator under the null hypothesis, $\underline{\hat{x}}_0 = A^+ \underline{y}$, with $A^+ = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$. The vector of residuals is obtained through Best Linear Unbiased Estimation (BLUE). The GLR test reads:

$$\begin{cases} \text{Accept } \mathcal{H}_0 \text{ if } T_q \leq k \\ \text{Reject } \mathcal{H}_0 \text{ if } T_q > k \end{cases} \tag{4}$$

where k is a positive scalar threshold.

The statistic $\underline{T}_q$ can be also written in alternative equivalent formulations as [19]:

$$\begin{aligned}
\underline{T}_q &= \underline{\hat{e}}_0^T Q_y^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_a^T Q_y^{-1} \underline{\hat{e}}_a = \|\underline{\hat{e}}_0\|^2_{Q_y^{-1}} - \|\underline{\hat{e}}_a\|^2_{Q_y^{-1}} \\
\underline{T}_q &= (\underline{\hat{y}}_0 - \underline{\hat{y}}_a)^T Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_a) \\
\underline{T}_q &= \underline{\hat{\nabla}}^T C^T P_A^{\perp T} Q_y^{-1} P_A^{\perp} C \underline{\hat{\nabla}} \\
\underline{T}_q &= \underline{\hat{\nabla}}^T Q_{\hat{\nabla}}^{-1} \underline{\hat{\nabla}}
\end{aligned} \tag{5}$$

where $\underline{\hat{e}}_a = \underline{y} - A\underline{\hat{x}}_a - C\underline{\hat{\nabla}}$ is the vector of observation residuals computed considering the alternative hypothesis holding true. $\underline{\hat{\nabla}}$ is the BLUE of the bias ∇. $\underline{\hat{y}}_0 = A\underline{\hat{x}}_0$ and $\underline{\hat{y}}_a = A\underline{\hat{x}}_a + C\underline{\hat{\nabla}}$ are respectively the vectors of estimators for the means of the measurements in the null and alternative hypotheses. P_A is the projector onto the space $R(A)$ under the metric defined by Q_y^{-1}, i.e. $P_A = AA^+$. Its corresponding orthogonal projector is defined as $P_A^{\perp} = I - P_A$. the squared norm of a vector v in the metric defined by the positive matrix M, $\|v\|^2_M$, is defined as $\|v\|^2_M \equiv v^T M v$. As from the first expression of (5), the GLR test statistic $\underline{T}_q$ is in fact the measurement-residual squared norm separation between null and alternative hypothesis.

With reference to the statistic defined in (3) and (5) and that the noise of each measurement is Gaussian, it holds that:

$$\begin{aligned}
\mathcal{H}_0 :&\quad \underline{T}_q \sim \chi^2(q, 0) \\
\mathcal{H}_a :&\quad \underline{T}_q \sim \chi^2(q, \lambda)
\end{aligned} \tag{6}$$

where the non-centrality parameter λ is expressed as:

$$\begin{aligned}
&\lambda = \nabla^T Q_{\hat{\nabla}}^{-1} \nabla \quad \text{or} \\
&\lambda = \|P_A^{\perp} C \nabla\|^2_{Q_y^{-1}}
\end{aligned} \tag{7}$$

3.1 Most powerful property

The test in (4), i.e. the GLR test, is the best test in the sense that it is UMPI (Uniform Most Powerful Invariant). By this we mean that it is Uniformly Most Powerful among the class of Invariant tests. For a detailed explanation of the meaning of these terms refer to [1]. With respect to the formulation of the testing problem we consider in this study, we define a test ϕ as Uniformly Most Powerful, with size α, if ϕ has size (significance level) α and if any other possible test ϕ^* with the same size has power $\gamma^*(\nabla) \leq \gamma(\nabla)$, for any ∇ and any x, where γ is the power of test ϕ. The power of the test (γ) is the probability of correctly rejecting the null hypothesis, when the alternative holds true. For invariance of a testing procedure under a group G of invertible functions g, the result of the test does not change if we observe the original observables $\underline{y}$ transformed through a function $g \in G$. This means that $\phi(g(\underline{y})) = \phi(\underline{y})$. For the proof of UMPI property see [1]. Finally note that this property holds for the case of 'binary' testing; i.e. testing a single alternative hypothesis against the null, as specified in 2.

3.2 Special form — Case $q = 1$

As a special case, when only a single satellite fault is considered, we can form an $m \times 1$ matrix C (a vector therefore) in the model in (2). In this case ∇ is reduced to a simple scalar and $q = 1$. The test $\underline{T}_{q=1}$ becomes:

$$\underline{T}_{q=1} = \frac{\underline{\hat{\nabla}}^2}{\sigma^2_{\hat{\nabla}}}$$

such that the random variable:

$$\underline{w} = \frac{\underline{\hat{\nabla}}}{\sigma_{\hat{\nabla}}} \tag{8}$$

is normally distributed, with:

$$\begin{aligned}
\mathcal{H}_0 :&\quad \underline{w} \sim N(0, 1) \\
\mathcal{H}_a :&\quad \underline{w} \sim N(\nabla w, 1)
\end{aligned} \tag{9}$$

and $\nabla w = \frac{\nabla}{\sigma_{\hat{\nabla}}}$. The random variable $\underline{w}$ is commonly referred to as the w-test statistic.

In case C is chosen as a canonical unit vector of the space R^m, and m alternative hypotheses are defined, one for each of the m dimensions (for instance, $C_{y_1} = [1\ 0\ \cdots\ 0]^T$ in case the first measurement y_1 is considered to be possibly faulty), we have the so-called data snooping; that means each measurement is checked individually. In case of data snooping, and furthermore when Q_y is diagonal, the w-test has the simple formula:

$$\underline{w}_i = \frac{\hat{e}_i}{\sigma_{\hat{e}_i}} \quad \text{for} \quad i = 1, 2, \ldots, m \tag{10}$$

where i refers to the presumed faulty measurement.

4 Solution Separation test

The Solution Separation test as first presented in [16] and adopted in the ARAIM algorithm [6, 10] employs as test statistics the difference between the all-in-view solution ($\mathcal{H}_0$) and the solution obtained excluding the measurements that possibly include threats ($\mathcal{H}_a$):

$$\underline{T}_{SS} = \underline{\hat{\nabla} \hat{x}} = \underline{\hat{x}}_0 - \underline{\hat{x}}_a \tag{11}$$

where $\underline{\hat{x}}_0$ is the solution under the null hypothesis while $\underline{\hat{x}}_a$ is the solution computed under the alternative hypothesis. Note here the use of double hat in $\underline{\hat{\nabla} \hat{x}}$, since it is the difference between the estimators of the expectations of the solutions under null and alternative hypotheses. The full vector is considered (or only a subvector of components of interest for integrity, but for simplicity of notation we consider the full vector $\underline{\hat{\nabla} \hat{x}}$ in the following), and a threshold is set for each tested mode. The test is:

$$\begin{cases} \text{Accept } \mathcal{H}_0 \text{ if } |\hat{\nabla} \hat{x}_i(j)| \leq k_{i,j} \quad \forall i = 1, \ldots, N_a \text{ and } \forall j = 1, 2, 3 \\ \text{Reject } \mathcal{H}_0 \text{ if } \exists\ i, j : |\hat{\nabla} \hat{x}_i(j)| > k_{i,j} \end{cases} \tag{12}$$

where $\exists\ i, j$ means 'exists at least one pair of numbers i, j', N_a is the total number of alternative hypotheses considered and j indicates a position component (East, North or Up). The absolute values of $\underline{\hat{\nabla} \hat{x}}_i$ position components are tested because we consider here symmetric detection regions with respect to $\hat{\nabla} \hat{x}_i = 0$ (expected value for $\nabla = 0$). The use of this test was justified in [7] mainly on heuristic and empirical grounds.

In 3D positioning, three tests are run for each alternative hypothesis $\mathcal{H}_a$ tested: one for each component of the statistic $\underline{\hat{\nabla} \hat{x}}$. The statistic $\underline{\hat{\nabla} \hat{x}}$ has a multivariate normal distribution, and:

$$\begin{array}{ll} \mathcal{H}_0 : & \underline{\hat{\nabla} \hat{x}} \sim N(0, Q_{\hat{\nabla} \hat{x}}) \\ \mathcal{H}_a : & \underline{\hat{\nabla} \hat{x}} \sim N(\nabla \hat{x}, Q_{\hat{\nabla} \hat{x}}) \end{array} \tag{13}$$

with $Q_{\hat{\nabla} \hat{x}} = A^+ C Q_{\hat{\nabla}} C^T A^{+T}$. When adopting the test in (12), the significance (α) and the power (γ) of the test are not straightforward to compute, since they correspond to integrals of a multivariate normal distribution over a box. In the ARAIM algorithm the significance α (which is associated to the False Alarm rate P_{FA}) is computed simply by applying the Bonferroni approximation, i.e. upperbounding the probability of occurrence of any of two events by the sum of their individual probabilities, as if they were mutually exclusive. In practice, to compute the thresholds for the tests, the total requirement on P_{FA} is allocated between the vertical, as $P_{\text{FA}_{\text{ver}}}$, and horizontal directions as $P_{\text{FA}_{\text{hor}}}$, respectively, where:

$$P_{\text{FA}} = P_{\text{FA}_{\text{hor}}} + P_{\text{FA}_{\text{ver}}} \tag{14}$$

and the $P_{\text{FA}_{\text{hor}}}$ allocation is further split into two, for each horizontal component, yielding the thresholds:

$$\begin{array}{rcl} k_{\text{ver}} & = & \Phi^{-1}(P_{\text{FA}_{\text{ver}}}) \\ k_{\text{hor}_1} = k_{\text{hor}_2} & = & \Phi^{-1}(\frac{P_{\text{FA}_{\text{hor}}}}{2}) \end{array} \tag{15}$$

The Bonferroni approximation is sharp when the components of $\underline{\hat{\nabla} \hat{x}}$ are not highly correlated, but this is not always the case, especially for alternative hypotheses with a few degrees of freedom. In case of single

satellite fault ($q = 1$), the three components are fully correlated. This means that the full P_{FA} can be used to set the threshold of each of the three tests, i.e.:

$$\begin{array}{rcl} k_{\mathrm{ver}} & = & \Phi^{-1}(P_{\mathrm{FA}}) \\ k_{\mathrm{hor}_1} = k_{\mathrm{hor}_2} & = & \Phi^{-1}(P_{\mathrm{FA}}) \end{array} \tag{16}$$

5 Solution Separation and GLR test

In this section we derive the relationship between the GLR test and the SS test presented in the previous sections. First we describe the main relationships between the test statistics and their representation in the observation and position domains. Next we analyze in more detail these relationships, treating separately the one-dimensional threat and the multi-dimensional threat cases.

5.1 Test statistics in the observation and solution domains

The SS statistic can be expressed as:

$$\underline{\hat{\nabla}\hat{x}} = \underline{\hat{x}}_0 - \underline{\hat{x}}_a = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\underline{\hat{\nabla}} \tag{17}$$

from which we can write:

$$A\underline{\hat{\nabla}\hat{x}} = P_A C\underline{\hat{\nabla}} = P_A\underline{\hat{\nabla}y} \tag{18}$$

As shown in Equation 5 a general formulation of $\underline{T}_q$ is:

$$\underline{T}_q = \|P_A^{\perp} C\underline{\hat{\nabla}}\|^2_{Q_y^{-1}} = \|P_A^{\perp}\underline{\hat{\nabla}y}\|^2_{Q_y^{-1}} \tag{19}$$

Therefore we note that the realization of the test statistic T_q is the norm (in the metric defined by matrix Q_y^{-1}) of the projection of the estimator of the bias vector $\hat{\nabla}y = C\hat{\nabla}$ in the space perpendicular to $R(A)$, whereas the SS is directly related to the projection of the $\hat{\nabla}y$ in the space $R(A)$. This means that the linear function of the SS test statistic (18) and the GLR test statistic (19) are the two orthogonal components of the same vector $\hat{\nabla}y$.

From the previous equation we can further derive:

$$\|\underline{\hat{\nabla}\hat{x}}\|^2_{Q_{\hat{x}_0}^{-1}} = \|P_A C\underline{\hat{\nabla}}\|^2_{Q_y^{-1}} \tag{20}$$

since $Q_{\hat{x}_0}^{-1} = A^T Q_y^{-1} A$. Therefore, the Pythagoras relation holds, and we can write the following relation between the GLR test statistic and the norm of the SS:

$$\|C\underline{\hat{\nabla}}\|^2_{Q_y^{-1}} = \|P_A^{\perp} C\underline{\hat{\nabla}}\|^2_{Q_y^{-1}} + \|P_A C\underline{\hat{\nabla}}\|^2_{Q_y^{-1}} = \underline{T}_q + \|\underline{\hat{\nabla}\hat{x}}\|^2_{Q_{\hat{x}_0}^{-1}} \tag{21}$$

This represents the central relation between the estimators of the biases in the observation and position domains, and can be visualized as in Figure 1. $\underline{\hat{\nabla}y}$ is obtained by projecting $\underline{y}$ onto $R(A\ C)$ to get $\underline{\hat{y}}_a$ and then decomposing $\underline{\hat{y}}_a$ in $\underline{\hat{\nabla}y} + A\underline{\hat{x}}_a$. Therefore the procedure to obtain the test statistics is: project $\underline{y}$ onto $R(A\ C)$ to get $\underline{\hat{\nabla}y} + A\underline{\hat{x}}_a$, then project $\underline{\hat{\nabla}y}$ onto $R(A)$ to obtain $A\underline{\hat{\nabla}\hat{x}}$ (representation through matrix A of the SS statistic $\underline{\hat{\nabla}\hat{x}}$) and onto $R(A)^{\perp}$ to obtain $P_A^{\perp}\underline{\hat{\nabla}y}$ whose norm is the $\underline{T}_q$ test statistic. The process is as follows (18, 19):

$$\begin{array}{rcl} \underline{\hat{\nabla}y} + A\underline{\hat{x}}_a & = & \underline{\hat{y}}_a = P_{A\ C}\underline{y} \\ A\underline{\hat{\nabla}\hat{x}} & = & P_A\underline{\hat{\nabla}y} \\ \underline{T}_q & = & \|P_A^{\perp}\underline{\hat{\nabla}y}\|^2_{Q_y^{-1}} \end{array} \tag{22}$$

where $P_{A\ C}$ is the projector onto the space $R(A\ C)$ under the metric defined by Q_y^{-1}.

The relation described by (21) is not the only one relating the two test statistics. In fact $\underline{\hat{\nabla}y}$ is constrained to lie on $R(C)$; if $R(C)$ spans any of the base vectors of $R(A)^{\perp}$, or in case $q > n$, to the same $\underline{\hat{\nabla}\hat{x}}$ may correspond multiple values of $\underline{T}_q$; otherwise, to $\underline{\hat{\nabla}\hat{x}}$ corresponds only one value of $\underline{T}_q$.

This second relation can also be visualized in Figure 1. In this case $R(A\ C)$ is the full R^m space, so $y \equiv P_{AC}y = \hat{y}_a$, but $\hat{\nabla}y$ lies on $R(C)$ so that $A\hat{\nabla}\hat{x}$ and $\sqrt{T_q}$ are fully determined as its two orthogonal components.

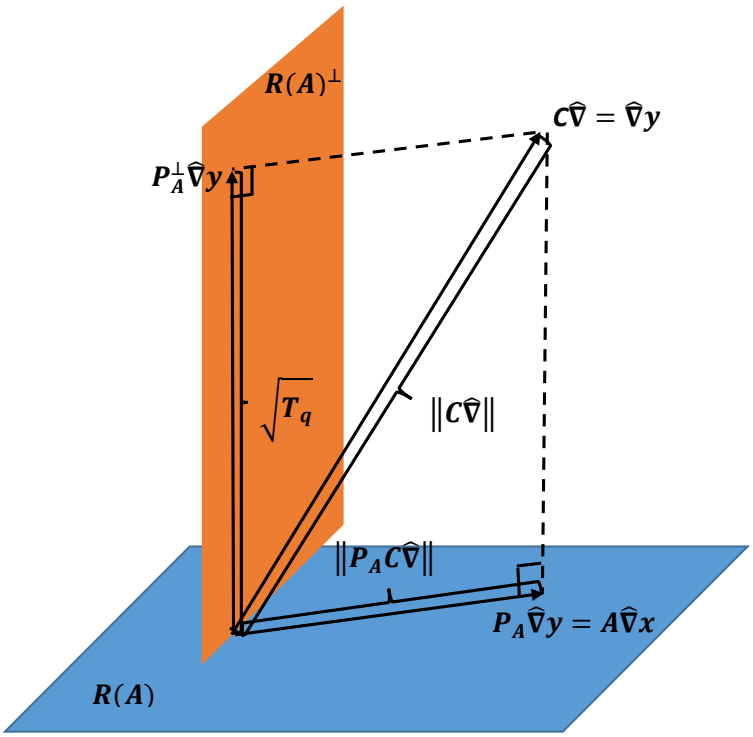


Figure 1: Geometrical visualization of T_q, $P_A C\hat{\underline{\nabla}}$ and $C\hat{\underline{\nabla}}$.

5.2 Biases in the observation and solution domains

The same relation as (21) holds between the *true* unknown biases in the observation and position domains, and was developed when defining the concepts of Internal and External Reliability of the test [19]. Internal Reliability relates to the power of the test to detect a bias in the observations, whereas the External Reliability relates to the effect an undetected bias has on the position solution. In particular, assuming a certain anomaly is present, causing a bias $C\nabla$ in the measurements, the corresponding effect on the position solution will be, as in (17), $\nabla\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\nabla$ (note the notation now with a single hat). With reference to [19], the following quantities can be also defined:

$$\lambda_{\hat{x}} = \nabla\hat{x}^T Q_{\hat{x}_0}^{-1} \nabla\hat{x} = \|A\nabla\hat{x}\|^2_{Q_y^{-1}} = \|P_A \nabla y\|^2_{Q_y^{-1}} \tag{23}$$

$$\lambda_y = \nabla y^T Q_y^{-1} \nabla y = \|\nabla y\|^2_{Q_y^{-1}} \tag{24}$$

with λ defined in (7). These quantities are visualized in Figure 2, which is the equivalent of Figure 1 for the *actual* error quantities. They allow writing in a simple way the relation between Internal and External Reliability. λ is in fact the non-centrality parameter of the distribution of the statistics $\underline{T}_q$ under the alternative hypothesis (7) and measures the incidence of the error on the test statistic (i.e. detectability of the fault). λ_y relates to the size of the detectable bias in the observation domain (Internal Reliability) while $\lambda_{\hat{x}}$ measures the effect on the position solution (External Reliability, i.e. effect of undetected fault). A similar visualization can be made for the quantities in (21). As a general relation, equivalent of (21), it holds:

$$\lambda_y = \lambda + \lambda_{\hat{x}} \tag{25}$$

This is the main relationship between Internal and External Reliability (λ_y and $\lambda_{\hat{x}}$).

5.3 Practical differences

We consider here two cases, i.e. one-dimensional and multi-dimensional anomalies.

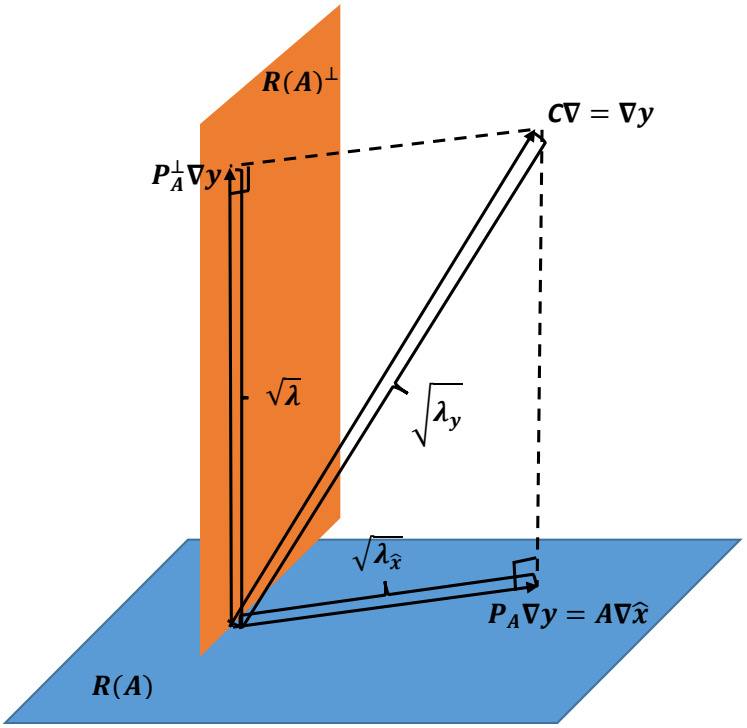


Figure 2: Geometrical visualization of λ, $\lambda_{\hat{x}}$ and λ_y.

- In case $q = 1$ (that is the case of $\underline{T}_q$ reduced to the w-test statistic) $\hat{\underline{\nabla}}$ is a scalar. Combining (19) and (20) we have:

$$\frac{\underline{T}_1}{\|\hat{\underline{x}}_0 - \hat{\underline{x}}_a\|^2_{Q^{-1}_{\hat{x}_0}}} = \frac{\|P_A^{\perp} C\hat{\underline{\nabla}}\|^2_{Q_y^{-1}}}{\|P_A C\hat{\underline{\nabla}}\|^2_{Q_y^{-1}}} = \frac{\|P_A^{\perp} C\|^2_{Q_y^{-1}}}{\|P_A C\|^2_{Q_y^{-1}}} \tag{26}$$

  Therefore in this case, as long as the denominator $\|P_A C\|^2_{Q_y^{-1}} \neq 0$, $\underline{T}_q$ and the norm of the SS are directly proportional, the proportionality constant being a function of the geometry matrices.

  Since the direction of $\hat{\underline{x}}_0 - \hat{\underline{x}}_a = \hat{\underline{\nabla}}\hat{x}$ is determined by C under given geometry, as can be seen in Equation 17, there is a univocal correspondence between a value of the statistic $\underline{T}_q$ and the SS $\hat{\underline{\nabla}}\hat{x}$, except for the sign ($\pm\hat{\underline{\nabla}}\hat{x}$ yield the same $\underline{T}_1$).

  If in the SS approach a threshold is set for each component of $\hat{\underline{\nabla}}\hat{x}$, only one of the n components actually constrains the SS, the tightest, since the direction of this vector is known.

  an upperbound to the test statistic $\underline{T}_q$ defines a hyper-ellipsoid in the space $R(C)$. The bias size $\hat{\underline{\nabla}}$ has q dimensions, whereas in the position domain $\hat{\underline{\nabla}}\hat{x}$ has n dimensions. Thus, in the comparison between $\underline{T}_q$ and $\hat{\underline{\nabla}}\hat{x}$ the dimensions q and n (n being the dimensions of $R(A)$ and also of the position solution) play important roles.

  Here we consider furthermore that the q-dimensional ellipsoidal constraint lies on $R(C)$, but we are fundamentally interested in its projection on $R(A)$. The projection of $R(C)$ on $R(A)$ is given by the matrix $P_A C = A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C$ (in the metric defined by Q_y^{-1}). Depending on whether $R(C)$ is perpendicular to $R(A)$ or not, the matrix $A^T Q_y^{-1} C$ has rank smaller than or equal to both q and n, $r = \text{rank}(A^T Q_y^{-1} C) \leq \min[q, n]$. r defines the number of dimensions on which a direct correspondence links the constraint (boundaries) in $R(A)$ given by the SS thresholds and the constraint in $R(C)$ given by the $\underline{T}_q$ threshold.

  In general, an upper-bound on $\underline{T}_q$ determines a q-dimensional ellipsoid in $R(C)$. Its projection on $R(A)$ will be an r-dimensional ellipsoid. Conversely, when a threshold is set in the SS approach for each component of $\hat{\underline{\nabla}}\hat{x}$, this determines a hyper-rectangular constraint in the position domain, to which corresponds a hyper-parallelogram in $R(A)$. Its planes will intersect the r-dimensional subspace

projection of $R(C)$ on $R(A)$, generating either a closed polyhedron or an open figure. The planes generate constraints in r dimensions. Therefore, they will create a polyhedron in the q-dimensional $R(C)$ only if $r \geq q$, otherwise the constraint does not constitute a full bound in the range bias $\hat{\underline{\nabla}}$ domain (and therefore there is no finite threshold for $\underline{T}_q$).

5.4 Comparison of the two methods

Since the SS test explicitly neglects the anomalies that have no effect in the position domain, we can expect that the power saved from neglecting those anomalies will be gained for the detection of actually dangerous anomalies. Even though it is possible that some gain is obtained (though difficult to quantify), there is still no evidence that the shape of the detection region is optimal from the integrity point of view.

When comparing SS and GLR tests [11], one can conclude that:

- In case of $q = 1$, i.e. a one-dimensional threat, SS and GLR tests are equivalent, except for very rare degenerate cases. Testing on each of the SS component is practically equivalent to test only one of them, the one creating the tightest bound, since the bias has only one dimension. The only cases in which SS and GLR tests can lead to different results occur when the vector C is perpendicular to the space $R(A)$, in which cases the SS would always accept the null hypothesis whereas the GLR w-test can still reject it. However, cases of exact perpendicularity are in practice very rare.

- In case of $q > 1$, SS and GLR tests may lead to different results. The main difference between the two approaches lies in the fact that the SS statistics refers only to the effect of possible outliers in the position domain. Hence, the outliers or faults that have no influence on the position solution will be completely neglected. This difference is especially acute when C spans any of the base vectors of the space perpendicular to $R(A)$, but also when $q > n$. In these cases biases that have no impact on the position solution can theoretically grow indefinitely in some 'directions' without being detected by the SS, while they are detected by the GLR. In the other cases the difference between the two tests lies in the shape of the detection region, which is ellipsoidal (for instance in the position bias domain) in the GLR case while bounded by pairs of parallel planes in the case of the SS.

6 Testing for hazardous biases with a new testing procedure

6.1 Hazardous bias size

We propose to apply statistical hypothesis testing for the detection of anomalies, starting off with differently defined hypotheses. The hypotheses previously introduced are basically $\mathcal{H}_0 : \nabla = 0$ and $\mathcal{H}_a : \nabla \neq 0$. We propose to test the null hypothesis $\mathcal{H}_0 : \nabla = 0$ versus the alternative $\mathcal{H}_a : \nabla \in \mathcal{B}$, where $\mathcal{B}$ represents the set of *hazardous* bias sizes. Therefore we start from:

$$\begin{array}{ll} \mathcal{H}_0 : & \underline{y} = Ax + \underline{e} \\ \mathcal{H}_a : & \underline{y} = Ax + C\nabla + \underline{e} \quad \nabla \in \mathcal{B} \end{array} \tag{27}$$

and we write the Generalized Likelihood Ratio (GLR) test as:

$$\text{Reject } \mathcal{H}_0 \text{ if } \ \text{GLR} \equiv \frac{\max f_{\underline{y}}(y|\mathcal{H}_a)}{\max f_{\underline{y}}(y|\mathcal{H}_0)} = \frac{\max f_{\underline{y}}(y|\nabla \in \mathcal{B})}{\max f_{\underline{y}}(y|\nabla = 0)} > k \tag{28}$$

where $f_{\underline{y}}(y)$ stands for the Probability Density Function (PDF) of $\underline{y}$. This yields the following test statistic:

$$\underline{T}' = \hat{\underline{e}}_0^T Q_y \hat{\underline{e}}_0 - \min_{\hat{\nabla} \in \mathcal{B}} \hat{\underline{e}}_a^T Q_y^{-1} \hat{\underline{e}}_a = \|\hat{\underline{e}}_0\|^2_{Q_y} - \min_{\hat{\nabla} \in \mathcal{B}} \|\hat{\underline{e}}_a\|^2_{Q_y^{-1}} \tag{29}$$

which can be reformulated in various ways — similar to (3) and (5). From estimation theory [1] we know that:

$$\arg\min_{\hat{\nabla}} \|\hat{\underline{e}}\|_{Q_y} = \hat{\nabla}_{\text{LS}} = (\bar{C}_y^T Q_y^{-1} \bar{C}_y)^{-1} \bar{C}_y^T Q_y^{-1} y \tag{30}$$

i.e. the Least Squares (LS) estimate for ∇, with $\bar{C}_y = P_A^{\perp} C$. Therefore we have:

$$\underline{T}' = \begin{cases} \|\underline{\hat{e}}_0\|^2_{Q_y^{-1}} - \|\underline{\hat{e}}_{\mathrm{LS}}\|^2_{Q_y^{-1}} & \text{if: } \hat{\nabla}_{\mathrm{LS}} \in \mathcal{B} \\ \|\underline{\hat{e}}_0\|^2_{Q_y^{-1}} - \min_{\hat{\nabla} \in \mathcal{B}} \|\underline{\hat{e}}\|^2_{Q_y^{-1}} & \text{if: } \hat{\nabla}_{\mathrm{LS}} \notin \mathcal{B} \end{cases} \tag{31}$$

In (31), one of the two minimums is reached at $\hat{e} = \hat{e}_{\mathrm{LS}}$ (i.e. $\hat{\nabla} = \hat{\nabla}_{\mathrm{LS}}$), and the other has to be searched inside or outside $\mathcal{B}$. In fact, assuming general regularity of the region $\mathcal{B}$, such minimum (the one different from $\|\underline{\hat{e}}_{\mathrm{LS}}\|^2_{Q_y}$) is reached for a $\hat{\nabla}$ corresponding to the point on the border of $\mathcal{B}$ that is the closest to $\hat{\nabla}_{\mathrm{LS}}$, in the metric defined by $Q_{\hat{\nabla}}$.

Figure 3 shows a geometric interpretation of the above test statistics, in the vector space $R(C)$. Two cases are shown, a case in which $\hat{\nabla}_{\mathrm{LS}}$ lies outside $\mathcal{B}$ (on the left), and a case in which $\hat{\nabla}_{\mathrm{LS}}$ lies inside $\mathcal{B}$ (on the right). In the space $R(C)$, $\|\underline{\hat{e}}_0\|_{Q_y^{-1}}$ corresponds to the distance between $\hat{\nabla}_{\mathrm{LS}}$ and the origin, $\|\underline{\hat{e}}_{\mathrm{LS}}\|_{Q_y^{-1}}$ is null and $\min_{\hat{\nabla} \in \mathcal{B}} \|\underline{\hat{e}}\|_{Q_y^{-1}}$ corresponds to the distance between $\hat{\nabla}_{\mathrm{LS}}$ and the border of $\mathcal{B}$. The test statistic is thus a measure of the 'distance' of the current observation from the null hypothesis and of the 'vicinity' to the alternative hypothesis. Therefore when $\hat{\nabla}_{\mathrm{LS}}$ lies outside $\mathcal{B}$, the distance from $\hat{\nabla}_a$ has to be subtracted.

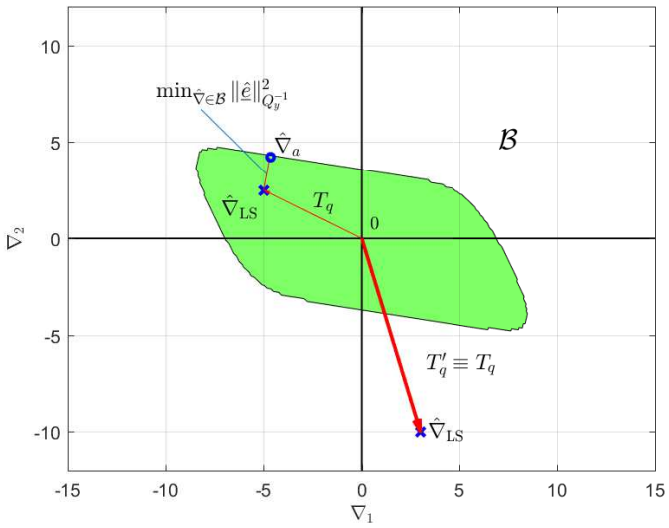


Figure 3: Graphical illustration of the optimal test for hazardous biases T'. Once $\hat{\nabla}_{\mathrm{LS}}$ is computed, we check if it lies in $\mathcal{B}$. If this is the case, T'_q coincides with the squared distance (in the metric defined by $Q_{\hat{\nabla}}$) of $\hat{\nabla}_{\mathrm{LS}}$ from the origin. Otherwise, T'_q is computed subtracting from the (squared) distance between $\hat{\nabla}_{\mathrm{LS}}$ and the origin the (squared) distance between $\hat{\nabla}_{\mathrm{LS}}$ and $\hat{\nabla}_a$, which is the point on the border of $\mathcal{B}$ closest to $\hat{\nabla}_{\mathrm{LS}}$. In fact the test statistic is a measure of the 'distance' of the current observation from the null hypothesis and of the 'vicinity' to the alternative hypothesis, therefore when $\hat{\nabla}_{\mathrm{LS}}$ lies outside $\mathcal{B}$, the distance from $\hat{\nabla}_a$ has to be subtracted.

The set of hazardous biases $\mathcal{B}$ in the observation domain shall be derived from the set of dangerous biases in the position domain, say $\tilde{\mathcal{B}}$. The set of hazardous biases is a function of the set of allowable positioning errors, say Ω_{AL}, which usually includes all the error vectors whose components are smaller than the Alert Limits (in absolute value), and of the allowable probability of Hazardous Misleading Information (P_{HMI}), say $\bar{P}_{\mathrm{HMI}}$. The P_{HMI} is the most important parameter of interest in integrity monitoring algorithms (e.g. RAIM), see [12, 13], and is defined as:

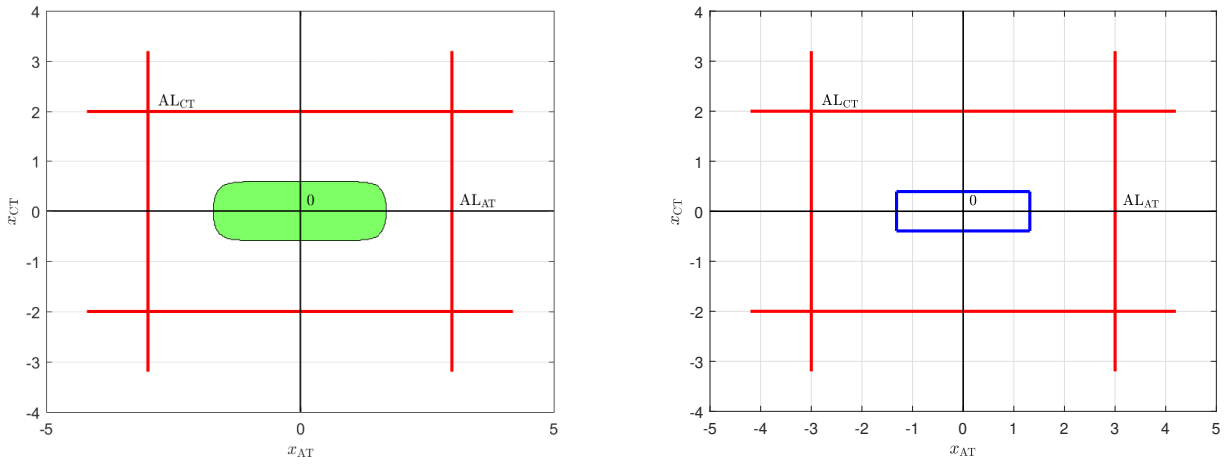
$$P_{\mathrm{HMI}} = P(\underline{\hat{x}} \notin \Omega_{\mathrm{AL}} \cap \text{ No Alarm }) \tag{32}$$

In case the fault detection algorithm in use is based on a single test, the issuing of an alarm can coincide with the rejection of the detection test, $T' > k$. Alert Limits and allowable $\bar{P}_{\mathrm{HMI}}$ are requirements defined by the authorities or convention, and are dependent on the navigation application at hand.

Thus $\tilde{\mathrm{B}}$ can be chosen as $\tilde{\mathcal{B}} = \{\nabla\hat{x} : P(\underline{\hat{x}} \notin \Omega_{\mathrm{AL}}|\mathcal{H}_a) > \bar{P}_{\mathrm{HMI}|\mathcal{H}_\mathrm{a}}\}$, where $\bar{P}_{\mathrm{HMI}|\mathcal{H}_\mathrm{a}}$ is the maximum allowed conditional P_{HMI}, possibly computed as $\bar{P}_{\mathrm{HMI}|\mathcal{H}_\mathrm{a}} = \bar{P}_{\mathrm{HMI}}/P_i$, with P_i the prior probability (assumed known) of occurrence of the ith alternative hypothesis. The distribution of $\hat{x}$ under any H_a is known as a function of $\nabla\hat{x}$, therefore $P(\underline{\hat{x}} \notin \Omega_{\mathrm{AL}}|\mathcal{H}_a) > \bar{P}_{\mathrm{PF}|\mathcal{H}_\mathrm{a}}$ can be computed (numerically) for any $\nabla\hat{x}$, as an integration of a multi-variate normal distribution over the region Ω_{AL}.

6.2 Test threshold

It is not possible to write in a closed form the probability distribution function of the test statistic defined in Equation (31). The test statistic expression is composed of three terms, i.e. $\|\underline{\hat{e}}_0\|^2_{Q_y^{-1}}$, $\|\underline{\hat{e}}_{\mathrm{LS}}\|^2_{Q_y^{-1}}$ and $\min_{\hat{\nabla}\in\mathcal{B}} \|\underline{\hat{e}}\|^2_{Q_y^{-1}}$. The first two are χ^2 distributed, but the PDF of the third cannot be derived analytically, and we have to resort to numerical computation. To simplify the computation of the test statistic PDF, we can approximate (conservatively) the region $\tilde{\mathrm{B}}$ to a region of rectangular shape, as shown in Figure 4. After this approximation, the numerical computation is simplified: $\min_{\hat{\nabla}\in\mathcal{B}} \|\underline{\hat{e}}\|^2_{Q_y^{-1}}$ can be found performing a linear search over the edge of $\tilde{\mathcal{B}}$, and the test threshold for a given significance α can be obtained by Monte Carlo integration.



(a) (b)

Figure 4: Areas of tolerable and hazardous biases in the position domain: (a) actual set, (b) approximated set.

6.3 Single Point Positioning (SPP) example

Consider the case of a two-dimensional fault ($q = 2$) for a SPP case with GPS+Galileo, with the satellite geometry shown in Figure 5. Let us assume that we are interested in monitoring the integrity in the horizontal directions, and Alert Limits are set to 3.5m for the North direction (say along-track when a vehicle is traveling towards North) and 2.5m for the East (cross-track) direction.

The $m \times 5$ design matrix A and pseudorange variance matrix Q_y are structured as

$$A = \begin{bmatrix} 1 & 0 & -u_1^T \\ \vdots & \vdots & \vdots \\ 1 & 0 & -u_{m_{\mathrm{GPS}}}^T \\ 0 & 1 & -u_{m_{\mathrm{GPS}}+1}^T \\ \vdots & \vdots & \vdots \\ 0 & 1 & -u_{m_{\mathrm{GPS}}+m_{\mathrm{Gal}}}^T \end{bmatrix}, \; Q_y = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2) \tag{33}$$

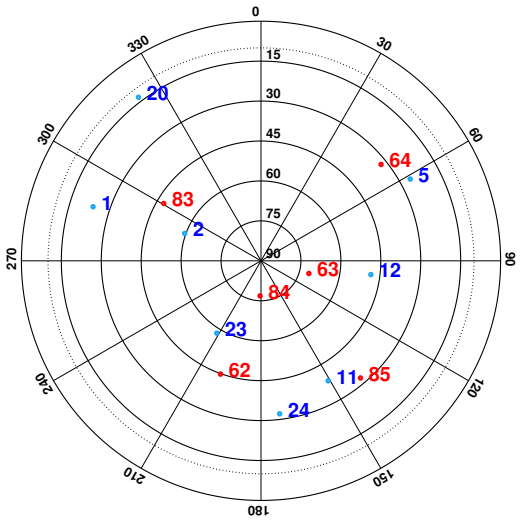


Figure 5: Skyplot GPS+Galileo. In blue the GPS satellites, in red the Galileo ones.

with u_i being the 3×1 ith receiver-satellite unit direction vector, in the North-East-Up frame, and m_{GPS} and m_{Gal} are the number of in-view satellites of the GPS and Galileo constellations, respectively. The unknown parameter vector consists of the receiver clock offsets of GPS and Galileo, and the increments to the three receiver position components. The stochastic model is based on ionosphere-free observations (from dual frequency L1 and L5 for GPS, E1 and E5a for Galileo), with the entries of the diagonal variance matrix constructed according to Tables 1 and 2, based on [8, 12, 17].

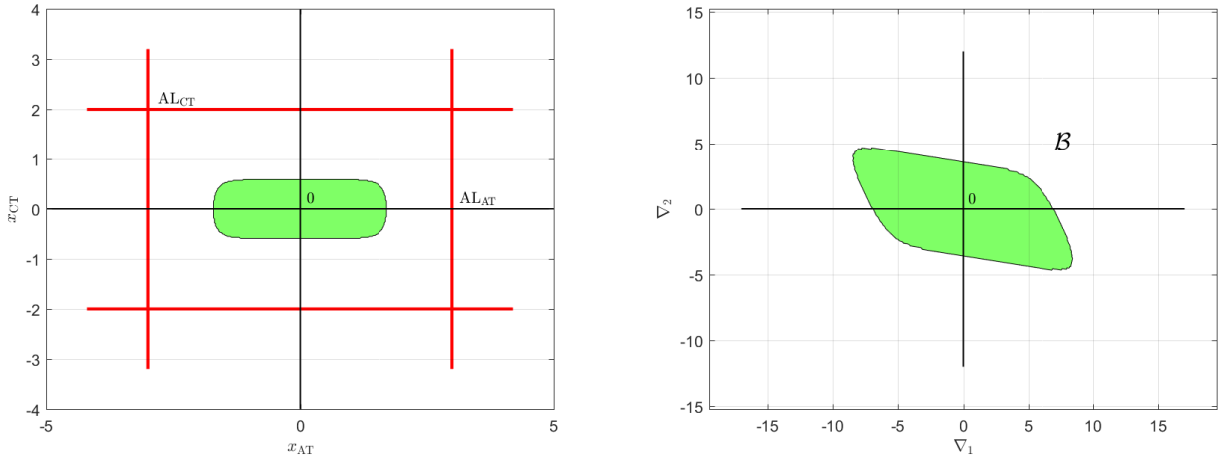
Table 1: Simulation parameters. ξ stands for the elevation of the satellite, in degrees.

Parameter	Description	Value
σ_i^{GPS}	GPS smoothed code total standard deviation used for integrity	$\sqrt{\sigma_{URA}^{GPS^2} + \sigma_{tropo}^2 + \sigma_{user}^{GPS^2}}$ m
σ_i^{Gal}	Galileo smoothed code total standard deviation used for integrity	$\sqrt{\sigma_{URA}^{Gal^2} + \sigma_{tropo}^2 + \sigma_{user}^{Gal^2}}$ m
σ_{URA}^{GPS}	GPS SV clock and orbit error used for integrity (URA)	0.75 m
σ_{URA}^{Gal}	Galileo SV clock and orbit error used for integrity (URA)	0.957 m
$\sigma_{tropo,i}$	Residual tropospheric error, for satellite i	$0.12\frac{1.001}{(0.002001+\sin^2\xi)^{1/2}}$ m
$\sigma_{MP,i}^{GPS}$	GPS smoothed code multipath	$0.13 + 0.53e^{-\xi/10}$ m
$\sigma_{noise,i}$	Smoothed code receiver noise	$0.15 + 0.43e^{-\xi/6.9}$ m
$\sigma_{user,i}^{GPS}$	GPS smoothed code total noise generated at user level	$\sqrt{\frac{f_{L1}^4+f_{L5}^4}{(f_{L1}^2-f_{L5}^2)^2}}\sqrt{\sigma_{MP}^2+\sigma_{noise}^2}$ m
$\sigma_{user,i}^{Gal}$	Galileo smoothed code total noise generated at user level	See Table 2

As $Q_{\hat{x}_0}$ is known, we can determine (numerically) the region that contains all the biases $\nabla\hat{x}_0$ that are assumed not dangerous, as the corresponding probability of positioning failure P_{PF} is within the requirement. Figure 6a shows in green the region of allowable position biases; the white area outside the green area is $\mathcal{B}$, the region of *hazardous* biases. Since the fault is two-dimensional in this example, to this area corresponds a closed region in the fault parameters space spanned by ∇_1 and ∇_2 (the two components of ∇, i.e. the parameters that define the anomaly), which is shown in Figure 6b.

Table 2: Galileo Elevation Dependent Signal in Space (SIS) user error.

Elevation	σ^{Gal}_{user}	Elevation	σ^{Gal}_{user}
5°	0.4529 m	50°	0.2359 m
10°	0.3553 m	55°	0.2339 m
15°	0.3063 m	60°	0.2302 m
20°	0.2638 m	65°	0.2295 m
25°	0.2593 m	70°	0.2278 m
30°	0.2555 m	75°	0.2297 m
35°	0.2504 m	80°	0.2310 m
40°	0.2438 m	85°	0.2274 m
45°	0.2396 m	90°	0.2277 m



(a) (b)

Figure 6: Areas of tolerable and hazardous biases in (a) the position domain and (b) the observation domain.

7 Simulations

We make use of the same SPP model described in Section 6.3, with the satellite geometry shown in Figure 5 — 8 GPS satellites and 6 Galileo satellites are visible. For this particular snapshot configuration, we assume an anomaly to be *always* present, affecting a pair of measurements (satellites). We simulate eight different scenarios: eight different pairs of satellites failing at the same time. The bias size of each affected measurement is first randomized, with values uniformly distributed between 0 m and 6 m (see Tables 3 and 4) and between 3 m and 10 m (Table 5), and then kept fixed at an arbitrary value of 4 m (Table 6). For each of the eight scenarios, we simulate measurements for 10^5 epochs. We then apply the three different testing procedures, the GLR test, the SS test and the test proposed in this paper ('New test').

Tables 3 to 6 show the results of all the simulations (eight pairs of satellites, two choices of random bias sizes), in terms of detection rate of the tests and empirical P_{HMI} (HMI rate). Tables 3 and 4 shows results for biases between 0 and 6 m but with two different values of the significance level α, i.e., taking α= 0.05 and 0.01, which are popular values of α. Tables 5 and 6 show two other cases of bias range, but limiting α to 0.01 to show the effect of the bias value. One should note that the most important parameter for integrity monitoring purpose is the rate of HMIs (the P_{HMI}). A lower figure for this parameter indicates a better performance of the test. In case different tests deliver the same HMI rate, it is then preferable a test that produces the lowest number of 'false' detections possible (as they could turn into false alarms, and consequently lower positioning continuity). A 'false' detection is defined here as a detection that is not justified by an error in the position that exceeds the Alert Limits (while it is declared by the presence of biases in the measurements).

Figure 7 shows the distribution of position estimator for three cases (three pairs of satellites failing): in

each case, the three testing methods (GLR shown in the sub-figures a, d, g, SS shown in the sub-figures b, e, h and the New test shown in the sub-figures c, f, i) are compared. 10^4 epochs are shown in each plot. The bias bin 3-10 m and the α of 0.01 were used for the plots. Rejected position solutions are shown as yellow/brown dots, accepted solutions are shown as blue crosses. As observed from Figure 7, the SS test has shown larger differences to the GLR and the New test. The New test appears to show similar results as the GLR test.

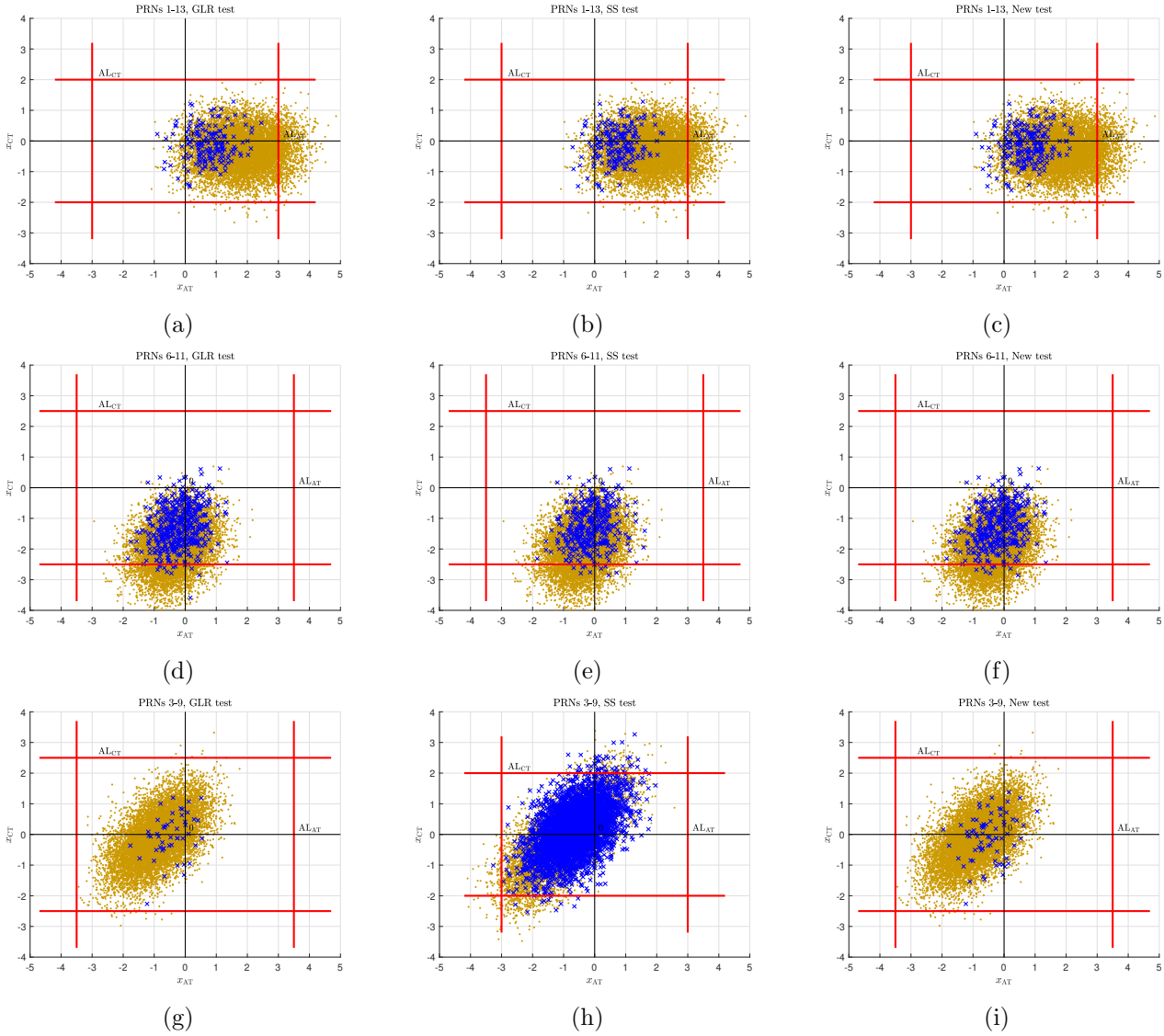


Figure 7: Positioning results from simulation of anomalies in different pair of observations, after application of the three testing methods, i.e., the GLR (a ,d, g), the SS (b, e, h) and the New test (c, f, i). The distribution of the position estimator is shown (10^4 epochs in each plot). Rejected position solutions are shown as yellow/brown dots, accepted solutions are shown as blue crosses. In the top panel the anomalous observations are from PRNs 1-2, in the middle panel from PRNs 6-11 and in the bottom panel from PRNs 3-9.

8 Discussion of the results

From the results in Tables 3 to 6, the following conclusions can be drawn, which are strictly based on the alternative hypothese considered in this study.

- All testing procedures are effective in detecting the anomalies, and their performance are comparable.

Table 3: Results of the different testing procedures in terms of detection rate and HMI rate for different scenarios. Bias sizes in two observations at a time are *randomized between* 0 *and* $6m$*, the significance* α *is chosen at* 0.05. 10^5 samples were simulated.

Scenario		Detection %			HMI %		
		GLR	SS	New test	GLR	SS	New test
1) Bias PRNs 1 & 13	$\alpha = 0.05$	79.82	74.87	71.55	< 0.01	< 0.01	< 0.01
2) Bias PRNs 6 & 11	$\alpha = 0.05$	64.99	64.18	66.83	0.08	0.07	0.05
3) Bias PRNs 3 & 9	$\alpha = 0.05$	79.86	22.46	76.91	< 0.01	0.04	< 0.01
4) Bias PRNs 5 & 6	$\alpha = 0.05$	76.65	76.68	78.00	0.01	0.01	0.01
5) Bias PRNs 2 & 10	$\alpha = 0.05$	77.24	73.08	77.59	0.01	0.01	0.01
6) Bias PRNs 8 & 9	$\alpha = 0.05$	74.04	72.20	74.80	0.54	0.54	0.50
7) Bias PRNs 13 & 14	$\alpha = 0.05$	74.68	47.02	69.23	0.01	0.02	0.01
8) Bias PRNs 4 & 8	$\alpha = 0.05$	73.91	72.59	74.54	0.77	0.73	0.70

Table 4: Results of the different testing procedures in terms of detection rate and HMI rate for different scenarios. Bias sizes in two observations at a time are *randomized between* 0 *and* $6m$*, the significance* α *is chosen at* 0.01. 10^5 samples were simulated.

Scenario		Detection %			HMI %		
		GLR	SS	New test	GLR	SS	New test
1) Bias PRNs 1 & 13	$\alpha = 0.01$	67.69	62.70	59.05	< 0.01	< 0.01	< 0.01
2) Bias PRNs 6 & 11	$\alpha = 0.01$	48.34	47.14	50.17	0.19	0.18	0.18
3) Bias PRNs 3 & 9	$\alpha = 0.01$	68.95	10.98	66.24	0.02	0.06	0.02
4) Bias PRNs 5 & 6	$\alpha = 0.01$	65.11	64.12	66.63	0.01	0.01	0.01
5) Bias PRNs 2 & 10	$\alpha = 0.01$	65.60	59.78	66.15	0.02	0.03	0.02
6) Bias PRNs 8 & 9	$\alpha = 0.01$	60.49	58.77	60.98	1.37	1.31	1.31
7) Bias PRNs 13 & 14	$\alpha = 0.01$	61.36	33.27	55.82	0.02	0.03	0.02
8) Bias PRNs 4 & 8	$\alpha = 0.01$	60.13	58.34	60.66	1.97	1.94	1.89

Table 5: Results of the different testing procedures in terms of detection rate and HMI rate for different scenarios. Bias sizes in two observations at a time are *randomized between* 3 *and* $10m$*, the significance* α *is chosen at* 0.01. 10^5 samples were simulated.

Scenario		Detection %			HMI %		
		GLR	SS	New test	GLR	SS	New test
1) Bias PRNs 1 & 13	$\alpha = 0.01$	99.37	99.26	98.76	0	0	0
2) Bias PRNs 6 & 11	$\alpha = 0.01$	94.99	95.79	95.79	0.13	0.11	0.09
3) Bias PRNs 3 & 9	$\alpha = 0.01$	99.60	30.92	99.33	0	0.06	0
4) Bias PRNs 5 & 6	$\alpha = 0.01$	99.16	99.12	99.33	< 0.01	< 0.01	< 0.01
5) Bias PRNs 2 & 10	$\alpha = 0.01$	98.85	98.74	99.14	< 0.01	0.01	< 0.01
6) Bias PRNs 8 & 9	$\alpha = 0.01$	97.88	98.45	98.02	0.47	0.34	0.43
7) Bias PRNs 13 & 14	$\alpha = 0.01$	98.38	78.41	97.72	< 0.01	0.02	< 0.01
8) Bias PRNs 4 & 8	$\alpha = 0.01$	97.06	97.65	97.17	0.90	0.71	0.86

Greater differences are registered between the GLR and the SS tests, whereas the performance of the GLR and the New test are similar.

Table 6: Results of the different testing procedures in terms of detection rate and HMI rate for different scenarios. Bias sizes in two observations at a time are $4m$ *in size, the significance* α *is chosen at* 0.01. 10^5 samples were simulated.

Scenario		Detection %			HMI %		
		GLR	SS	New test	GLR	SS	New test
1) Bias PRNs 1 & 13	$\alpha = 0.01$	94.9	94.9	90.4	< 0.01	< 0.01	< 0.01
2) Bias PRNs 6 & 11	$\alpha = 0.01$	65.8	73.5	69.9	0.44	0.36	0.40
3) Bias PRNs 3 & 9	$\alpha = 0.01$	96.8	3.5	94.0	< 0.01	< 0.01	< 0.01
4) Bias PRNs 5 & 6	$\alpha = 0.01$	93.4	93.1	94.7	0.03	0.03	0.02
5) Bias PRNs 2 & 10	$\alpha = 0.01$	89.6	91.2	92.5	0.05	0.04	0.03
6) Bias PRNs 8 & 9	$\alpha = 0.01$	81.7	87.4	82.7	3.66	2.56	3.49
7) Bias PRNs 13 & 14	$\alpha = 0.01$	86.3	45.9	81.6	0.01	0.05	0.01
8) Bias PRNs 4 & 8	$\alpha = 0.01$	75.4	80.7	76.2	5.81	4.53	5.62

- In all scenarios the New test shows the same or lower HMI rate than the GLR test, and, in case of same HMI rate, mostly a lower Alarm rate. This behaviour is expected since it is supposed to give less weight to anomalies with small bias size, therefore reducing corresponding 'false' alarms.
- With reference to the eight scenarios analyzed, scenarios 2, 6, 8 (as declared in Tables 3–6) show a better performance of the SS test over the GLR test, i.e. lower HMI rate. Scenarios 3, 7 (and 4 when larger biases are injected, Table 5) show a better performance of the GLR test over the SS test. Scenario 5 shows better results of the GLR test in case of random bias sizes (Tables 3 to 5) but better results of the SS test in case of fixed bias sizes (Table 5); however, in this scenario, the SS test is further outperformed by the New test.
- Overall, in the scenarios considered, the GLR test performed better than the SS test in terms of relative performance (HMI percentage reduction), whereas in absolute terms (e.g. summing up the HMI rates over all the scenarios considered) the SS test slightly outperformed the GLR. This is due to the fact that the SS test shows a better performance in the scenarios that are more critical for integrity (detection of satellite faults that lead to larger number of HMIs). However, there is no clear evidence to prefer the SS test over the GLR test, and more testing in different scenarios would be needed to come to more definitive conclusions.
- The improved performance of the New test over the GLR comes at the cost of increased complexity and computational load. For computation of 10^5 samples of one satellite pair, e.g. PRNs 3 & 9 with α of 0.01 and biases randomized between 3 and 10 m, the New method takes significantly more than double the computation time of the GLR method.

9 Conclusions

In this contribution the two statistical tests most commonly used in RAIM algorithms, the Generalized Likelihood Ratio (GLR) test and the Solution Separation (SS) test, were described, compared and discussed. Differences arise in testing multi-dimensional anomalies, whereas in the case of one-dimensional anomalies the two methods are equivalent. A new test was proposed, with a modification of the GLR test to address only faults that represent a threat to the solution integrity. The different testing procedures were then tested through simulations in multi-constellation SPP scenarios. The simulations results based on the tested scenarios led to the following conclusions:

- GLR and SS tests perform differently in each of the scenarios analysed, with the SS showing better results in the more integrity-critical scenarios, and the GLR showing more balanced results across all the scenarios. Overall, however, there is no evidence of better performance of one test over the other.

- The alternative procedure proposed (New test) shows a better performance than the standard GLR test, however this comes at the price of increased complexity and computational load.

The results obtained suggest that both GLR and SS tests, the most commonly used in RAIM algorithms, are not optimal for integrity monitoring purpose, since they were designed to detect any (generic) anomaly affecting the GNSS observations. A margin of improvement exists for the design of a monitoring technique that specifically addresses those anomalies that represent a threat to positioning.

Acknowledgements

This study is supported by the Australian Research Council project number DP170103341.

References

[1] Arnold, S. (1981). *The theory of linear models and multivariate analysis*, volume 2. Wiley, New York.

[2] Baarda, W. (1967). Statistical concepts in geodesy. Technical Report 4, Netherlands Geodetic Commission, Publ. on Geodesy, New Series.

[3] Baarda, W. (1968). A testing procedure for use in geodetic networks. netherlands geodetic commission, publ. on geodesy, new series. 2(5).

[4] Blanch, J., Walter, T., and Enge, P. (2013). Results on the optimal detection statistic for integrity monitoring. In *Proceedings of the 2013 International Technical Meeting of The Institute of Navigation*, pages 262–273.

[5] Blanch, J., Walter, T., and Enge, P. (2017). Theoretical results on the optimal detection statistics for autonomous integrity monitoring. *Navigation: Journal of The Institute of Navigation*, 64(1):123–137.

[6] Blanch, J., Walter, T., Enge, P., Lee, Y., Pervan, B., Rippl, M., Spletter, A., and Kropp, V. (2015). Baseline advanced RAIM user algorithm and possible improvements. *IEEE Transactions on Aerospace and Electronic Systems*, 51(1):713–732.

[7] Brown, R. G. and McBurney, P. W. (1987). Self-contained GPS integrity integrity checks using maximum solution separation as the test statistic. In *Proceedings of the First Technical Meeting of the Satellite Division of the Institute of Navigation, Colorado Springs, Colorado, USA*, pages 263–268.

[8] Choi, M., Blanch, J., Akos, D., Heng, L., Gao, G., Walter, T., and Enge, P. (2011). Demonstrations of multi-constellation advanced RAIM for vertical guidance using GPS and GLONASS signals. In *Proceedings ION GNSS 2011, Portland, OR*, pages 3227–3234.

[9] El-Mowafy, A. and Yang, C. (2016). Limited sensitivity analysis of araim availability for lpv-200 over australia using real data. *Advances in Space Research*, 57(2):659–670.

[10] Ene, A., Blanch, J., and Powell, J. (2007). Fault detection and elimination for Galileo-GPS vertical guidance. In *Proceedings of the Institute of Navigation's National Technical Meeting*. Citeseer.

[11] Imparato, D. (2014). Detecting multi-dimensional threats: A comparison of solution separation test and uniformly most powerful invariant test. In *Proceedings of the European navigation conference (ENC)-GNSS 2014*, pages 1–13. Nederlands, Instituut voor Navigatie.

[12] Imparato, D. (2016). *GNSS based Receiver Autonomous Integrity Monitoring for Aircraft Navigation*. Ph.D. Thesis, TU Delft.

[13] Joerger, M., Stevanovic, S., Langel, S., and Pervan, B. (2016). Integrity risk minimisation in RAIM part 1: Optimal detector design. *UK Journal of Navigation*, 69(03):449–467.

[14] Misra, P. and Enge, P. (2006). *Global Positioning System Signals, Measurements, and Performance*. Ganga-Jamuna Press, Second Edition, P.O. Box 692 Lincoln, Massachusetts 01773.

[15] Neyman, J. and Pearson, E. (1992). *On the problem of the most efficient tests of statistical hypotheses*. Springer.

[16] Parkinson, B. W. and Axelrad, P. (1987). A basis for the development of operational algorithms for simplified gps integrity checking. In *Institute of Navigation, Technical Meeting*, pages 269–276.

[17] Rippl, M., Martini, I., Belabbas, B., and Meurer, M. (2014). ARAIM operational performance tested in flight. In *Proceedings of the 2014 International Technical Meeting of The Institute of Navigation (ITM 2014)*, pages 601–615.

[18] Teunissen, P. J. G. (1990). An integrity and quality control procedure for use in multi sensor integration. In *Proceedings of the 3rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1990), September 19 - 21, 1990, Colorado Spring, CO*, pages 513–522.

[19] Teunissen, P. J. G. (2000). *Testing theory: an introduction*. Delft University Press. Series on Mathematical Geodesy and Positioning.